

Bambey, Doris; Reinhold, Anke; Rittberger, Marc

Pädagogik und Erziehungswissenschaft

Neuroth, Heike [Hrsg.]; Strathmann, Stefan [Hrsg.]; Oßwald, Achim [Hrsg.]; Scheffel, Regine [Hrsg.]; Klump, Jens [Hrsg.]; Ludwig, Jens [Hrsg.]: Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Hülsbusch, W 2012, S. 111-135



Quellenangabe/ Reference:

Bambey, Doris; Reinhold, Anke; Rittberger, Marc: Pädagogik und Erziehungswissenschaft - In: Neuroth, Heike [Hrsg.]; Strathmann, Stefan [Hrsg.]; Oßwald, Achim [Hrsg.]; Scheffel, Regine [Hrsg.]; Klump, Jens [Hrsg.]; Ludwig, Jens [Hrsg.]: Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Hülsbusch, W 2012, S. 111-135 - URN: urn:nbn:de:0111-opus-54383 - DOI: 10.25656/01:5438

<https://nbn-resolving.org/urn:nbn:de:0111-opus-54383>

<https://doi.org/10.25656/01:5438>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc-sa/3.0/de/deed> - Sie dürfen das Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen: Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses Werk bzw. der Inhalt darf nicht für kommerzielle Zwecke verwendet werden. Die neu entstandenen Werke bzw. Inhalte dürfen nur unter Verwendung von Lizenzbedingungen weitergegeben werden, die mit denen dieses Lizenzvertrages identisch oder vergleichbar sind.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nc-sa/3.0/de/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public and alter, transform or change this work as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work. If you alter, transform, or change this work in any way, you may distribute the resulting work only under this or a comparable license.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de



Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

Kapitel 6

Pädagogik und

Erziehungswissenschaft



Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft



Universitätsverlag Göttingen

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme



Universitätsverlag Göttingen

vwh
Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft



Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme
hg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump,
Jens Ludwig
im Rahmen des Kooperationsverbundes nestor – Kompetenznetzwerk Langzeitarchivierung
und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Lektorat) und Sonja Neweling (Redaktion) sowie Martina Kerzel und Lajos Herpay (Gestaltung und Montage) für ihre unverzichtbare Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Die Inhalte des Buches stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/bestandsaufnahme/>)
sowie über den Göttinger Universitätskatalog
(<http://www.sub.uni-goettingen.de>) zur Verfügung.
Die URN lautet: <http://nbn-resolving.de/urn:nbn:de:0008-2012031401>.

Die digitale Version dieses Werkes steht unter einer Creative Commons Namensnennung-Nicht-kommerziell-Weitergabe unter gleichen Bedingungen 3.0 Unported Lizenz.



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2012
<http://www.vwh-verlag.de>
In Kooperation mit dem Universitätsverlag Göttingen.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide
Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-86488-008-7

6 Pädagogik und Erziehungswissenschaft

Doris Bambey, Anke Reinhold, Marc Rittberger

6.1 Einführung in den Forschungsbereich

Die *Pädagogik* beschreibt die Lehre, die Theorie und die Wissenschaft von der Erziehung und Bildung von Kindern und Erwachsenen in Familie, Schule, Freizeit und Beruf.¹ Der Terminus *Erziehungswissenschaft* wird häufig mit dem Begriff der Pädagogik gleichgesetzt, ist aber insbesondere im Bereich der empirisch orientierten Forschung verbreiteter – weil hier die methodische Nähe zu sozialwissenschaftlich-empirischen Methoden stärker zum Ausdruck kommt.² Das Fach Erziehungswissenschaft gliedert sich in mindestens 25 Subdisziplinen und Fachrichtungen.³ Mit ca. 165.000 Studierenden⁴ stellt es heute in Deutschland das fünftgrößte Universitätsfach dar.

Die Pädagogik bzw. die Erziehungswissenschaft erfüllt eine Doppelrolle: Sie erforscht Bildungs- und Erziehungszusammenhänge und betreibt zugleich sozialwissenschaftlich-empirische Forschung über das Bildungswesen, d.h. *empirische Bildungsforschung*, die wissenschaftliche Informationen erhebt und auswertet, „die eine rationale Begründung bildungspraktischer und bildungspolitischer Entscheidungen ermöglichen“⁵. Von besonderem Interesse sind u.a. Fragen nach dem Verlauf von Bildungsprozessen sowie dem Erwerb von Qualifikationen und Kompetenzen im Bildungssystem.⁶

1 Vgl. Lenzen (2005), S. 1105.

2 Neuerdings findet auch der Begriff der *Bildungswissenschaft* stärkere Verbreitung im Forschungsbereich (vgl. Sailer (2007), S. 127).

3 Vgl. ZfE (2010).

4 Das BMBF fasst in den aktuellen Grund- und Strukturdaten erstmals Lehrerausbildung und Erziehungswissenschaft zu einer gemeinsamen Fachrichtung zusammen (vgl. BMBF (2007/2008), S. 38).

5 Tippelt; Schmidt (2010), S. 9.

6 Vgl. Gräsel (2011), S. 13.

Die wichtigsten Förderer von Bildungsforschungsprojekten in Deutschland sind die DFG, das BMBF und die Europäische Union. Diese großen Förderer finanzieren ca. zwei von fünf Projekten. Die restlichen Projekte werden hauptsächlich von Stiftungen, Verbänden und Unternehmen finanziell gefördert.⁷ In der Bildungsforschung spielen daneben auch ländergeförderte Projekte eine wichtige Rolle. Aktuelle Themen – beispielsweise des BMBF-Rahmenprogramms zur Förderung der Empirischen Bildungsforschung⁸ – sind u.a. die *Lebenslaufforschung*, die *Lehr-Lern-Forschung*, die *Kompetenzdiagnostik* bzw. *Kompetenzentwicklung* in Vorschule, Schule und Hochschule sowie die Etablierung des Nationalen Bildungspanels (NEPS⁹) (s. Kap. 6.2). Weitere Schwerpunkte sind Steuerungsfragen, Disparitäten im Bildungssystem (ethnisch, sozial, geschlechtsspezifisch und regional) sowie die Professionalisierung des pädagogischen Personals.¹⁰

Bekannte Forschungsprojekte der empirischen Bildungsforschung sind die großen vergleichenden Schulleistungsstudien wie z.B. PISA¹¹, TIMSS¹² und IGLU/PIRLS¹³. Diese lösten Mitte der 1990er-Jahre die sogenannte *zweite empirische Wende* in der Bildungsforschung aus.¹⁴ Eine zunehmend evidenzbasierte Politik verstärkt den Bedarf nach steuerungsorientierten und vergleichenden Evaluationen. Neuere Beispiele dafür sind – auch mit jeweils internationalem Fokus – die Projekte DESI¹⁵ und PIAAC¹⁶. Das Sozio-Ökonomische Panel (SOEP) – bisher die einzige genuin nationale Panelstudie – stellt nur in geringem Maße Daten zu Bildungsverläufen und Kompetenzentwicklung zur Verfügung.¹⁷ Diese Tatsache forcierte maßgeblich die Etablierung von NEPS.

7 Vgl. Huth (2012), S. 87.

8 Vgl. BMBF (2011f).

9 Vgl. NEPS (2011).

10 Vgl. BMBF (2007), S. 10ff.

11 Die Studie wird seit dem Jahr 2000 in dreijährigem Turnus durchgeführt.

12 Datenerhebungen: 1996, 1999, 2003, 2007 und 2011.

13 Die Studie wird seit dem Jahr 2001 in fünfjährigem Turnus durchgeführt.

14 Vgl. Watermann (2007), S. 214.

15 Laufzeit: 2001–2008.

16 Laufzeit: 2008–2013.

17 Vgl. Blossfeld (2009), S. 5f.

Die Bildungsforschung hat einen dezidiert inter- und multidisziplinären Charakter.¹⁸ Je nach Forschungsfeld finden psychologische, soziologische, ökonomische oder fachdidaktische Perspektiven Berücksichtigung.¹⁹ So weisen die *Arbeitsmarktforschung* oder die *Schulleistungsforschung* sachliche Überschneidungen mit den Wirtschafts- und Sozialwissenschaften auf. Weitere Berührungspunkte – insbesondere in der Lehr-Lern-Forschung – ergeben sich mit der *Pädagogischen Psychologie* sowie der Geschichtswissenschaft im Bereich der *Historischen Bildungsforschung*. Nicht zuletzt implizieren die Fachdidaktiken Bezüge zu nahezu allen Fächern.

6.2 Kooperative Strukturen

Institutsübergreifende Zusammenarbeit

Die interdisziplinäre Ausrichtung der Bildungsforschung und die starke Auffächerung ihrer Forschungsfelder sorgen für eine ausgeprägte Methodenvielfalt und dezentral vorliegende Datenbestände. Überproportional häufig werden kleinere Projekte an Hochschulen, Forschungseinrichtungen oder in Verbünden durchgeführt.²⁰ Dabei sind Projektlaufzeiten von ein bis zwei Jahren eher die Regel als die Ausnahme. Der Anteil kooperativ durchgeführter Projekte an der Gesamtforschung steigt zwar stetig an, spielt jedoch mit derzeit ca. 11% (noch) eine nur untergeordnete Rolle.²¹

Generell erschweren die verteilten Strukturen und die heterogene Datenlage die institutsübergreifende Zusammenarbeit. Bisher gibt es noch keine Mindestanforderungen und Standards für die Veröffentlichung und Sicherung von Forschungsdaten, und gerade kleinere Einrichtungen und Projekte verfügen selten über eine zentrale Anlaufstelle für nachhaltiges Datenmanagement. Eine institutsübergreifende Kooperation erfolgt in der Regel nur dann, wenn die forschungsstrategische Notwendigkeit dazu besteht. Beispiele hierfür sind die großen Schulleistungsstudien oder

18 Vgl. Tippelt/Schmidt (2010), S. 10.

19 Vgl. Prenzel (2005), S. 12.

20 „Bildungsforschung besteht zu 80 Prozent aus kleineren Projekten mit nur einem oder zwei Mitarbeitern“ (Schulzeck 2008, S. 41).

21 Vgl. Schulzeck (2008), S. 40.

die Panelstudien. Hier ist die Politik besonders daran interessiert, Sekundäranalysen zur steuerungorientierten Evaluation der Bildungsqualität zu initiieren – was das Vorhandensein eines kooperativ abgestimmten Datenmanagements voraussetzt.

Beispielhaft für eine solche prototypische Kooperation ist das NEPS-Projekt. Die Längsschnittstudie zu zentralen Bildungsprozessen und -verläufen über die gesamte Lebensspanne wird durch ein interdisziplinäres Konsortium von Forschungsinstituten und Forschergruppen durchgeführt.²² Die verschiedenen Projektsäulen werden im ersten Schritt – je nach inhaltlicher oder methodischer Expertise der beteiligten Partner – dezentral bearbeitet. Im zweiten Schritt werden die generierten Forschungsdaten in ein kooperativ betriebenes Data Warehouse überführt. Diese zentrale Hosting-Struktur bietet anschließend die Grundlage für ein mehrstufiges Konzept des Datenzugriffs sowie die Weiterentwicklung von Metadatenschemata in der Domäne (s. Kap. 6.3).

Projekte wie NEPS stärken einerseits die Entwicklung von Datenaustauschverfahren und die Abstimmung von Standards. Andererseits ist die Strategie des Datenmanagements im Projekt nicht in erster Linie auf die Weiterentwicklung von generischen Metadatenstandards ausgerichtet. Im Vordergrund steht vielmehr der unmittelbar zielorientierte und effiziente Projektverlauf. Zudem werden die Hosting-Strukturen (bisher) ausschließlich für NEPS-Daten zur Verfügung gestellt, womit sich der Trend zur Dezentralisierung des Datenzugangs tendenziell verstärkt.

Dezentrale oder zentralisierte LZA-Dienste

Bisher gibt es für die Bildungsforschung noch keinen zentralen LZA-Dienst für Forschungsprimärdaten. Entsprechend den disziplinübergreifenden Forschungsbereichen sind für die halbamtlichen Daten zwei relevante Datenprovider zu nennen: das *International Dataservice Center (IDSC)* des Forschungsinstituts für die Zukunft der Arbeit (IZA) und das *German Microdata Lab (GML)* der GESIS. Weitere bildungsrelevante Forschungsdaten sind derzeit über verschiedene Forschungsdatenzentren (FDZen) mit

22 Die Projektpartner sind insgesamt 15 namhafte Forschungsinstitute sowie Lehrstühle von 15 Universitäten in Deutschland. Die Koordination erfolgt über das Institut für bildungs-wissenschaftliche Längsschnittforschung (INBIL) an der Otto-Friedrich-Universität in Bamberg.

jeweils eigenen Zugriffsplattformen verteilt.²³ Eine zentrale Meta-Plattform für einen fokussierten Zugriff auf Forschungsdaten der Bildungsforschung (d.h. ein Datenservicezentrum) existiert jedoch bisher nicht. In größeren kooperativen Projekten werden z.T. eigene Serviceinfrastrukturen geschaffen – wie etwa bei NEPS – oder es werden vorhandene genutzt, wie z.B. die des *Instituts für die Qualitätsentwicklung im Bildungswesen* (IQB) für die großen Schulleistungsstudien der Bundesländer (s. Kap. 6.3).

Nachdem mit Blick auf den Zugang zu anonymisierten Mikrodaten verschiedene Forschungsdatenzentren in Deutschland eingerichtet wurden²⁴, richtet sich der Fokus nunmehr auf die nichtamtlichen, wissenschaftsgetragenen Forschungsdaten bzw. auf eine Verkopplung dieser Daten mit denjenigen der amtlichen Domäne. Gerade die projekt- und wissenschaftsbasierte Forschung ist durch eine stark verteilte Datenproduktion und eine erhebliche Methodenvielfalt charakterisiert und weist eine – auch technologisch bedingte – dynamische Entwicklung im Bereich neuer Analyse- und Erhebungsverfahren auf, wie z.B. bei NEPS²⁵.

Diese methodische Vielfalt wird jedoch kaum oder nur teilweise aufgefangen durch Verfahren vergleichsfördernder, standardisierter Datendokumentation und -information seitens der wissenschaftlichen Datenproduzenten selbst. Die Forschungslandschaft ist insgesamt stark geprägt durch Angebote von Einrichtungen und Projekten der Bildungsforschung, die sich inhaltlich und formal stark unterscheiden (s. Kap. 6.3). Somit besteht die Situation, dass es viele Einzelangebote gibt, es aber letztlich an einer zentralen Einrichtung mangelt, die unter der Perspektive der Langzeitarchivierung (LZA) die Koordination von Datendokumentation/-management, Datenzugriff/-sicherheit und Standardentwicklung betreibt.

23 Dabei handelt es sich um Forschungsdatenzentren an folgenden Instituten: FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB), FDZ des Instituts für Qualitätsentwicklung im Bildungswesen (IQB), FDZ der Deutschen Rentenversicherung, FDZ im Bundesinstitut für Berufsbildung (BIBB), FDZ „Wahlen“ und ALLBUS (GESIS), das IDSC am IZA. Zudem sind bildungsrelevante Daten über das Statistische Bundesamt Deutschland sowie die Statistischen Ämter des Bundes und der Länder zugänglich.

24 Diese Entwicklung hat ihren Ausgangspunkt in der Einrichtung der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) im Jahr 1999 sowie der Gründung des Rats für Sozial- und Wirtschaftsdaten (RatSWD) im Jahr 2001 (vgl. RatSWD (2011b)).

25 Vgl. Rolf; Zwick; Wagner (2008), S. 631ff.

Im Rahmen des DFG-geförderten Projekts „Kooperative Langzeitarchivierung erziehungswissenschaftlicher e-Ressourcen im Rahmen von *kopal*²⁶“ (LZA-Pädagogik) werden derzeit erstmals ein prototypischer Archivierungs-Workflow sowie ein Kooperationsmodell für digitale Objekte erarbeitet, das Forschungsdaten der Domäne explizit einschließt. Das Projekt wird seit 2009 vom Deutschen Institut für Internationale Pädagogische Forschung (DIPF) in Zusammenarbeit mit der Deutschen Nationalbibliothek (DNB) durchgeführt. In enger Anbindung an das Projekt DP4lib²⁷ sollen erziehungswissenschaftliche Archivierungsobjekte (u.a. Digitalisate historischer pädagogischer Zeitschriften und Nachschlagewerke²⁸ sowie audiovisuelle Forschungsdaten²⁹) langzeitarchiviert werden. Das Projekt nutzt somit die Expertise verschiedener fachlicher Dienstleister und bibliothekarischer Einrichtungen innerhalb der bestehenden nationalen LZA-Strukturen.

Das vorrangige Ziel des Projekts ist, Bearbeitungsabläufe zur Langzeitarchivierung zu erarbeiten, die den gesamten LZA-Prozess – von der Objektauswahl und Qualitätssicherung über die Datenlieferung an u.a. die DNB bis zum Datenrücktransfer nach erfolgter Migration – umfassen (s. Abb. 1). Für die audiovisuellen Forschungsdaten stehen die Festlegung eines einheitlichen und offenen Datenformats, die Definition LZA-relevanter Metadaten sowie inhaltliche, rechtliche und technische Prüfroutinen im Vordergrund. Die Datensicherung im Sinne einer *bitstream preservation* ist langfristig – d.h. auf die von der DFG empfohlene Zeitperspektive von zehn Jahren³⁰ – angelegt, um dem Anspruch auf eine zukünftige Interpretierbarkeit der Daten gerecht zu werden.

26 Das BMBF-geförderte Projekt *kopal* (Laufzeit 2004–2007) wurde durchgeführt unter der Leitung der DNB und zusammen mit den Projektpartnern SUB Göttingen, der IBM Deutschland GmbH sowie dem Rechenzentrumspartner GWDG. Ziel war die Entwicklung eines Informationssystems zur kooperativen Langzeitarchivierung digitaler Objekte (vgl. *kopal* (2011)).

27 Das DFG-geförderte Projekt der DNB und der SUB Göttingen ist als Nachfolgeprojekt und zur Fortentwicklung von *kopal* angelegt.

28 Inhalte der Datenbank *Scripta Paedagogica Online* aus dem Bestand der Bibliothek für Bildungsgeschichtliche Forschung (BBF).

29 Datenbank „Audiovisuelle Medien für die Unterrichtsforschung“ im Rahmen des Open-Access-Dokumentenservers „*pedocs*“ (vgl. *peDOCS* (2011); Kap. 6.3).

30 Vgl. DFG (1998).

Die folgende Abbildung zeigt schematisch eine Workflow-Variante für Archivierungsobjekte aus einem Datenspeicher in das LZA-System, wie sie zurzeit für das Projekt DP4lib entwickelt wird. Das Modell sieht vor, die Archivierungsobjekte nach einer Dateiformatmigration sowohl an den externen Datenspeicher zur Nutzung als auch in den DNB-eigenen Archivspeicher zur weiteren LZA zurückzuliefern.

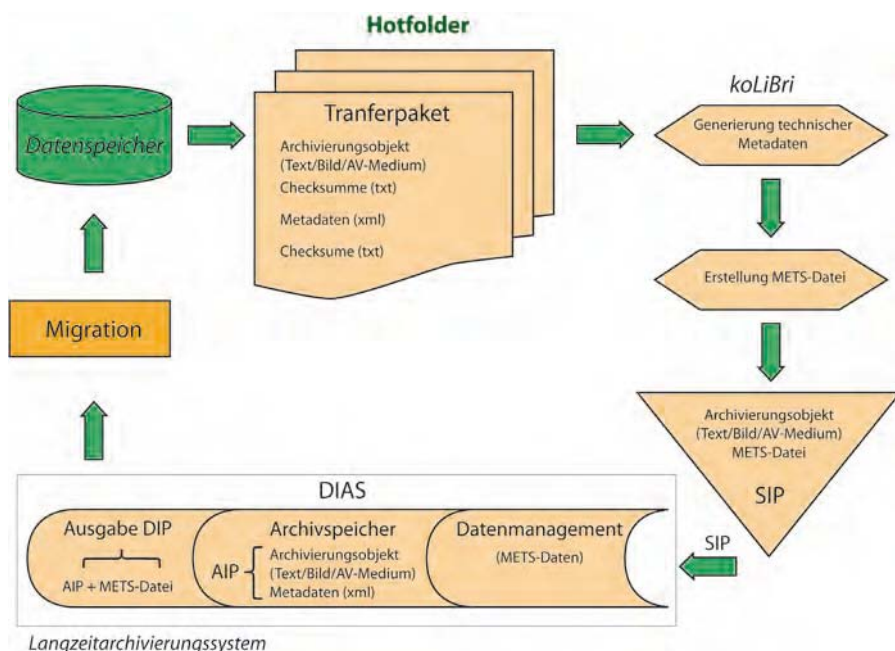


Abb. 1: Beispiel für eine Workflow-Variante für Archivierungsobjekte im Projekt DP4lib³¹

Insbesondere für den Objekttyp der audiovisuellen Forschungsdaten soll das Projekt LZA-Pädagogik eine Vorreiterrolle einnehmen. In diesem Sinne werden alle im Projekt erarbeiteten Materialien – wie z.B. methodische Leitfäden, Tools und Mustervereinbarungen, Anforderungen und Empfehlungen – dokumentiert, so dass eine Nachnutzung bzw. Adaption durch andere Institutionen und Disziplinen ermöglicht wird.

Nach wie vor werden Forschungsprimärdaten in der Bildungsforschung – insbesondere solche, die in kleineren Projekten anfallen – häufig bei

31 Vgl. Kreusch (2010).

den Forschern oder Forschergruppen selbst archiviert, d.h. auf Festplatten, Laufwerken oder portablen Speichermedien. Ein erster Vorstoß zur Verkopplung von Forschungsaktivitäten mit LZA-Infrastrukturen wird durch das DP4lib-Projekt für audiovisuelle Daten (AV-Daten) vorbereitet. Als eine zentrale Plattform für die koordinierte Aggregation von bildungsrelevanten Forschungsdaten in unmittelbarer Koppelung mit einem LZA-Workflow bietet sich das Fachportal Pädagogik³² an. So zielt das Projekt DP4lib darauf, im Rahmen des Fachportals Pädagogik audiovisuelle Daten aus verschiedenen Projekten des Forschungsbereichs *Unterrichtsqualität* prototypisch in einen kooperativen LZA-Workflow einzubinden.

6.3 Daten und Metadaten

Datenarten und Datenzugang in der Bildungsforschung

Für eine umfassende Bestandsaufnahme der in der Bildungsforschung anfallenden Forschungsprimärdaten bietet sich eine Unterscheidung in quantitative und qualitative Daten an.³³ In der erziehungswissenschaftlichen Bildungsforschung werden – abhängig von den Fragestellungen – unterschiedlich ausgeprägt quantitative oder qualitative Ansätze verfolgt bzw. aufeinander bezogen.³⁴ Nach wie vor sind jedoch quantitative Large-Scale-Ansätze mit hoher Reichweite für Öffentlichkeit und Politik deutlich visibler, wie es z.B. die großen Vergleichsstudien der Schulleistungsforschung demonstrieren. Jedoch lässt sich seit den 1970er-Jahren eine zunehmende nationale wie internationale Akzeptanz qualitativer Ansätze feststellen.³⁵ Häufig werden auch beide Ansätze innerhalb einzelner Studien kombiniert.³⁶

In den internationalen Vergleichsstudien der quantitativen Bildungsforschung erfolgt die Datengewinnung vorzugsweise durch objektive Tests, Skalen und Fragebogen.³⁷

32 Vgl. Fachportal Pädagogik (2011b).

33 Die Unterteilung des methodischen Spektrums in quantitative und qualitative Bildungsforschung findet sich beispielsweise bei Tippelt; Schmidt (2010).

34 Vgl. Tippelt; Schmidt (2010), S. 9f.

35 Vgl. Krüger (2000), S. 327.

36 Vgl. Friebertshäuser; Prengel (2003), S. 11.

37 Vgl. Seel (2010), S. 552ff.

„Kennzeichnend ist die Überführung verbaler oder visueller Informationen in Zahlenformate [...] und die Weiterverarbeitung dieser Zahlen. [...] Im Rahmen der Empirischen Bildungsforschung werden quantitative Verfahren häufig im Bereich der Unterrichtsforschung, Leistungs- und Kompetenztests eingesetzt.“³⁸

Die qualitative Bildungsforschung hingegen untersucht die Perspektiven der Akteure im Bildungsbereich möglichst „im Feld“. Als Materialbasis dienen dabei die Transkriptionen von Audio- oder Videoaufzeichnungen.³⁹ Derzeit besteht in der qualitativen Bildungsforschung jedoch eine klare Dominanz von Befragungen.⁴⁰ Diese methodische Unterscheidung ermöglicht die Typisierung der Daten in quantitative und qualitative Daten. Gleichzeitig ist jedoch zu beachten, dass bestimmte Datenarten, wie z.B. Fragebogen – je nach Erkenntnisinteresse – sowohl quantitativ als auch qualitativ ausgewertet werden können.

Die Sekundäranalyse von Forschungsdaten spielt für den gesamten Forschungsbereich bisher nur eine untergeordnete Rolle. Für insgesamt nur 14,7% aller Bildungsforschungsprojekte in Deutschland wird die Sekundäranalyse als Methode angewandt.⁴¹ Die Gründe dafür könnten einerseits darin liegen, dass die Forscher nicht wissen, wo und wie sie auf vorliegende Daten zugreifen können bzw. mit welchen Methoden eine Re- oder Sekundäranalyse durchführbar ist. Andererseits verbinden Forscher mit der Erhebung „eigener Daten“ gegebenenfalls auch die Erwartung, dass ihre Forschungsleistung eine höhere Anerkennung erfährt.

Zugang zu quantitativen Daten der Bildungsforschung

Bei den großen international angelegten Vergleichsstudien der empirischen Bildungsforschung fallen große Mengen an quantitativen Daten an, die für Re- und Sekundäranalysen relevant sind. Die Politik ist daran interessiert, einen inhaltlichen Mehrwert aus den steuerungrelevanten Evaluationen zu gewinnen. Zudem spielen Kostenaspekte eine Rolle. Am Beispiel von PISA, TIMSS und IGLU/PIRLS wird deutlich, dass die

38 Reinders; Gniewosz (2011), S. 121.

39 Vgl. Krüger (2000), S. 334ff.

40 Vgl. Schulzeck (2008), S. 51.

41 Davon sind ca. zwei Drittel als Befragungen und ca. ein Drittel als Aggregatdatenanalyse gekennzeichnet (vgl. Schulzeck (2008), S. 49).

Verfügbarkeit von Forschungsdaten im internationalen Kontext schon weit fortgeschritten ist. Hier sind die Datensätze sowie umfangreiches Dokumentationsmaterial frei im Internet zugänglich und ohne Antragsverfahren wiederverwertbar. Für PISA 2009⁴² sind beispielsweise die Fragebogen und Codebücher als PDF-Dokumente zugänglich, die Daten können als SAS und SPSS-Dateien heruntergeladen und verarbeitet werden. Zudem steht ein Auswahlmenü zur Verfügung, über das verschiedene Länder und Variablen kombiniert werden können. Auch für TIMSS und PIRLS ist diese Art des Datenzugangs⁴³ bereits verwirklicht.

Für den Zugang zu den Forschungsdaten der großen ländergetragenen Schulleistungstudien in Deutschland existiert seit 2007 das nach dem Königsteiner Schlüssel⁴⁴ durch die Bundesländer finanzierte FDZ am IQB⁴⁵ in Berlin. Nutzer können ausschließlich auf Antrag⁴⁶ und mit berechtigtem Interesse die nationalen Datensätze einsehen und für Re- und Sekundäranalysen verwenden. Bisher sind neun Studien beim IQB verfügbar.⁴⁷ Der Zugang zu den Daten erfolgt – je nach Sensibilität der Daten – entweder über ein *Scientific-Use-File* (in der Regel im SPSS-Format und als Leerdatensatz⁴⁸) oder über einen Gastwissenschaftler-Arbeitsplatz am IQB in Berlin. Zur Bearbeitung sensibler Daten besteht die Möglichkeit des Fernrechnens mithilfe der vom IDSC entwickelten Software *JoSuA* am IZA in Bonn. Ausschließlich über den Gastwissenschaftler-Arbeitsplatz besteht die Möglichkeit des Zugriffs auf die Länderkennungsvariablen.⁴⁹

42 Die aktuellen PISA-Daten sind im Internet erhältlich (vgl. PISA (2009). Für die Daten der PISA-Studien 2000, 2003 und 2006 vgl. PISA (2011)).

43 Vgl. TIMSS (2011).

44 „Der Königsteiner Schlüssel regelt die Aufteilung des Länderanteils bei gemeinsamen Finanzierungen. Er wird vor allem für die Aufteilung des Anteils der Länder an den Zuschüssen für die DFG, die MPG und die Einrichtungen der Leibniz-Gemeinschaft angewandt (§4 Abs. 1 AV-DFG, §4 Abs. 2 AV-MPG und §6 Abs. 1 AV-FE)“ (GWK (2008)).

45 Das Institut orientiert sich in seiner Arbeit an den „Kriterien des RatSWD für Forschungsdatenzentren“ (vgl. IQB (2011d)).

46 Nutzer können online den Antrag auf Datenzugang stellen (vgl. IQB (2011a)).

47 Dazu gehören ELEMENT, IGLU 2001, IGLU 2006, MARKUS, PISA 2000, PISA 2003, PISA 2006, die Pilotierungsstudie IQB Französisch und QuaSUM. In Vorbereitung sind die Studien DESI und LAU (vgl. IQB (2010)).

48 Der Nutzer hat Zugriff auf die verwendeten Variablen und die dazugehörigen Wertelabels, jedoch nicht auf die Länderkennung.

49 Vgl. Knigge (2011).

Für Studien, die über das IQB verfügbar gemacht werden sollen, liegen bereits Empfehlungen für die Datenabgabe an das Forschungsdatenzentrum vor: „Datensätze und Instrumente sollen nach den Beschlüssen der KMK jeweils ein Jahr nach dem Erscheinen des betreffenden Hauptberichts bzw. drei Jahre nach der Datenerhebung [...] übergeben werden“.⁵⁰ Die Daten werden in der Regel im SPSS-Format, die Dokumentationen in MS Office- oder PDF-Formaten übergeben. Allgemein lässt sich sagen, dass die formalen Anforderungen an die Daten gering sind. Forscher, die ihre Daten übergeben möchten, erhalten bei Bedarf Unterstützung, z.B. wenn unvollständige Dokumentationen vorliegen. Als konkrete Empfehlung und Orientierungshilfe für das zeitnahe Datenmanagement bereits während der Konzeptionsphase einer Studie nennt das IQB die Vorgaben des Projekts CESSDA⁵¹. Stanat/Döbert (2009) fordern darüber hinaus, dass insbesondere die Daten der Large-Scale-Studien möglichst schnell nach der Erhebung bereinigt und skaliert dem IQB übermittelt werden müssen. Zudem müsse die Transparenz des Antragsprozesses gewährleistet sein, d.h. die Forscher sollten konkrete Hinweise über genehmigte und abgelehnte Fragestellungen erhalten und inhaltsbezogene Einschränkungen sollten vermieden werden.⁵²

Trotz der Einbindung der Forschungsprojekte in einen politischen Auftrag blickt das IQB auf eine zunächst zögerliche Datenübergabe durch die Bildungsforscher zurück. Das IQB sah sich mit verschiedenen Einwänden konfrontiert und nennt diese in einem Erfahrungsbericht zur Datenübergabe: 1) Erhöhter Dokumentationsaufwand, 2) Verlust der exklusiven Rechte an den Daten (Balance zwischen persönlichem Aufwand und Nutzen) sowie 3) hoher administrativer Aufwand durch notwendige Reaktionen auf falsche Durchführung und Interpretation von Analysen. Zudem äußerten die beteiligten Bundesländer Bedenken bezüglich eines befürchteten Kontrollverlusts über die Analysen und Daten.⁵³

Im Ergebnis wurde ein äußerst restriktiver Datenzugang implementiert. Die beteiligten Datenproduzenten haben ein Vetorecht bezüglich der Sekundäranalyse der Daten, und für Länder-Vergleiche muss die

50 IQB (2011c).

51 Vgl. CESSDA (2011a).

52 Vgl. Stanat; Döbert (2009), S. 8ff.

53 Vgl. Knigge (2009).

Einwilligung der beteiligten Bundesländer eingeholt werden. Schließlich müssen zwei unabhängige Forscher jeden Antrag bezüglich der Umsetzung wissenschaftlicher Standards überprüfen. Auf der Webseite des IQB können alle Anträge eingesehen werden, die bisher zur Veröffentlichung ganz oder teilweise freigegeben wurden. Bisher wurden dort seit 2008 insgesamt 76 Anträge veröffentlicht.⁵⁴ Die PISA 2003-Datensätze wurden mit insgesamt ca. 36 Prozent der Anträge am stärksten nachgefragt; dahinter folgen PISA 2000 mit ca. 22 Prozent, PISA 2006 mit ca. 19 Prozent und die Datensätze der ELEMENT⁵⁵-Studie mit ca. 14 Prozent der Anträge auf Datenzugang.⁵⁶

Für die Bildungsforschung sind zudem die Informationsdienstleistungen der GESIS relevant. Neben dem Datenarchiv für Sozialwissenschaften, über das auch auf bildungsbezogene Forschungsarbeiten zugegriffen werden kann, existiert seit Juli 2010 die Registrierungsagentur da|ra. Dort können Forschungsdaten per DOI registriert und somit dauerhaft identifiziert werden (GESIS 2011).⁵⁷ Über diese Datenbank stehen den Nutzern unter dem Schlagwort Bildung derzeit Informationen zu 468 Studien⁵⁸ und den dazugehörigen Forschungsdaten zur Verfügung.

Ein weiteres Projekt, das die Bereitstellung von quantitativen Forschungsdaten der empirischen Bildungsforschung zum Ziel hat, ist die „Datenbank Schulqualität (DaQS)“⁵⁹ zu Instrumenten und Konstrukten der Schul- und Unterrichtsqualität mit dem Schwerpunkt Grund- und Sekundarschule. Die Instrumentendatenbank ermöglicht den Zugang zu den in den Studien des Forschungsbereichs eingesetzten Erhebungsinstrumenten und relevanten Kennzahlen. Die Nutzer können u.a. auf Studiensteckbriefe, Fragebogen, Skalen und Skalenhandbücher zugreifen. Zudem umfassen die – ohne Restriktionen verfügbaren – Recherchemöglichkeiten die Bereiche Theoriebildung, Untersuchungsplanung und

54 Vgl. IQB (2011b).

55 Erhebung zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin (Untersuchungen: 2003, 2004 und 2005).

56 Vgl. Knigge (2011).

57 Vgl. GESIS (2011c).

58 Stand: 19. April 2011 (vgl. GESIS (2011b)).

59 Das Projekt wird vom DIPF durchgeführt (Laufzeit: 2009–2011).

Operationalisierung.⁶⁰ Um die verschiedenen Studien auf einem Level abbilden zu können, wird zudem ein an gegebenen Standards orientiertes Metadatenschema für DaQs entwickelt. Die Datenbank steht seit März 2011 über das Fachportal Pädagogik online zur Verfügung.⁶¹

Auch das Deutsche Jugendinstitut (DJI) bearbeitet bildungsbezogene Themen, z.B. den Übergang vom Kindergarten in die Schule oder den Ausbau von Ganztagschulen. Daten aus DJI-Studien fließen u.a. in die nationale Bildungsberichterstattung⁶² ein. Eine *Online Survey*-Datenbank⁶³ ermöglicht den Zugriff auf Datensätze im SPSS-Format und Dokumentationen der Studien im PDF-Format; vergleichbare Variablen werden durch Querbezüge manuell gekennzeichnet. Die Datenbank bietet eine Volltextsuche in den Variablen und das Filtern nach Kategorien. Zudem können die Nutzer analog zu einer Sitemap in einer Datamap navigieren, um Studien und dazugehörige Datensätze aufzufinden.

Neben diesen Beispielen für den institutionalisierten Zugriff auf quantitative Forschungsdaten stellen auch einzelne Projekte mit Bildungsbezug ihre Daten über Web-Schnittstellen der wissenschaftlichen Öffentlichkeit zur Verfügung. Ein Beispiel ist *pairfam*⁶⁴, eine DFG-geförderte Längsschnittstudie zur Erforschung partnerschaftlicher und familialer Lebensformen in Deutschland. Seit 2010 existiert das unter dem Dach des RatSWD eingerichtete FDZ *pairfam*⁶⁵. Die Datensätze stehen auf Antrag als anonymisierte Scientific-Use-Files zur Verfügung und werden via CD-ROM an die Nutzer versandt. Die Befragungsdaten liegen in Stata-, SPSS- und SAS-Formaten vor; zudem kann der Nutzer auf Dokumentationsmaterial im PDF-Format zurückgreifen.

Bei NEPS (s. Kap. 6.1) – das bisher noch nicht als FDZ im Sinne des RatSWD fungiert – können die verschiedenen Nutzergruppen zukünftig über drei Zugangswege auf die Scientific-Use-Files (Stata-, SPSS-, SAS- oder R-Dateien) im Data Warehouse zugreifen: Download der

60 Mittelfristig soll z.B. ein Tool entwickelt werden, das die individuelle Zusammenstellung eigener Fragebogen aus einzelnen Fragen erlaubt.

61 Vgl. DaQS (2011).

62 Vgl. Bildungsbericht (2010).

63 Vgl. DJI (2011).

64 Partner des Projekts sind die Universität Bremen, die Technische Universität Chemnitz, die Universität Mannheim und die Ludwig-Maximilians-Universität München.

65 Vgl. *pairfam* (2011).

Forschungsdaten über die NEPS-Website, „On-Site“-Datenzugriff an der Universität Bamberg sowie Fernzugriff über die Technologie *RemoteNEPS*. Damit bringt NEPS hohe Anforderungen an eine koordinierte Aufbereitung, Zugänglichmachung und Vernetzung der Erhebungsdaten sowie an Datenschutzmaßnahmen mit sich – abhängig von den Analyseinteressen und Berechtigungen der jeweiligen Datennutzer.

Zugang zu qualitativen Daten der Bildungsforschung

Auch im Bereich der qualitativen Forschungsdaten wurden für die Bildungsforschung erste Infrastrukturen zur Bereitstellung entwickelt. In den vergangenen Jahren wurden – entweder an erziehungswissenschaftlichen Fachbereichen von Hochschulen oder auf Projektbasis – verschiedene Datensammlungen mit unterschiedlichen Zielsetzungen und Datenarten aufgebaut. Häufig basieren sie auf dem persönlichen Engagement der Verantwortlichen und unterscheiden sich hinsichtlich ihrer Professionalität und der Datenmenge. Daneben existiert für den Bereich der audiovisuellen Forschungsdaten am DIPF eine DFG-geförderte Infrastruktur, die bereits verschiedene Datenbestände erschlossen und datenbankgestützt gebündelt hat (s. Kap. 6.2).

In der empirischen Bildungsforschung werden häufig qualitative Ansätze verfolgt, um Teilaspekte von primär quantitativ-statistischen Untersuchungen zu bearbeiten. Ein Beispiel hierfür sind die TIMSS-Studien, bei denen begleitend zur Leistungsstanderhebung mithilfe von Befragungen der Unterricht in verschiedenen Kulturen mit Videokameras beobachtet wurde⁶⁶, oder die Videostudien im Kontext der Projekte DESI und VERA⁶⁷. Die wissenschaftliche Dokumentation und Evaluation erfolgte mithilfe der Methode der Unterrichtsvideografie. Beispiele für die Generierung audiovisueller Forschungsdaten in der Bildungsforschung sind

66 Vgl. Kiel (2010), S. 785.

67 Vgl. VERA (2011).

zudem die Projekte Pythagoras⁶⁸ und PERLE⁶⁹. Qualitative Videostudien haben insbesondere in der Unterrichtsforschung – auch in kleineren Bildungsforschungsprojekten – in den vergangenen Jahren an Bedeutung gewonnen. Zentrale Forschungsfragen betreffen die Unterrichtsqualität bzw. den Unterrichtserfolg. Neben den audiovisuellen Daten fallen auch weitere Datenarten wie Fragebogen, Interviewtranskripte, digitale Fotos (z.B. Tafelbilder) etc. an.

Für die Forschungsdaten des Projekts Pythagoras wird derzeit in Anbindung an das Fachportal Pädagogik die Datenbank „Audiovisuelle Medien für die Unterrichtsforschung“ zu Konstrukten der Schul- und Unterrichtsqualität aufgebaut (s. Kap. 6.2). Im Rahmen des DFG-geförderten Projekts werden der wissenschaftlichen Community Rohdaten in Form videobasierter Unterrichtsaufzeichnungen sowie umfangreiches Dokumentationsmaterial zur Verfügung gestellt. Insgesamt liegen ca. 360 Unterrichtsbeobachtungen, Lehrerinterviews und Aufzeichnungen tutorieller Situationen sowie zugehörige Transkripte vor; der Umfang des Datenvolumens wurde nach einer internen Auswertung des DIPF auf 550 bis 600 Gigabyte geschätzt. Wissenschaftliche Nutzer sollen diese Forschungsdaten zugangsbeschränkt als passwortgeschützte PDF-Dateien herunterladen und für Sekundäranalysen nutzen können. Auch die Forschungsdaten aus dem PERLE-Projekt sollen in dieser Form bereitgestellt werden.

Bereits abgeschlossen wurde das Projekt „Videodatenbank Schulunterricht in der DDR“⁷⁰. Unterrichtsaufzeichnungen aus der ehemaligen DDR wurden datenbanktechnisch und dokumentarisch aufbereitet und

68 Die schweizerisch-deutsche Studie wurde kooperativ durch das DIPF und das Pädagogische Institut der Universität Zürich durchgeführt. In 20 deutschen Klassen der 9. Jahrgangsstufe und 20 Schweizer Klassen der 8. Jahrgangsstufe wurde die Qualität des Mathematikunterrichts u.a. am Beispiel der Unterrichtseinheit „Satz des Pythagoras“ untersucht. In Deutschland war das Projekt eingebettet in das Schwerpunktprogramm „Bildungsqualität von Schule“ (2000–2006). Hintergrund der Untersuchung war die bessere Mathematik-Performance der Schweizer Schüler bei den PISA-Tests (vgl. Hugener; Pauli; Reusser (2006), S. 5).

69 In der Studie wird seit 2005 mit einem kombinierten Quer- und Längsschnittsdesign die Entwicklung von Grundschulkindern in Sachsen untersucht. Dafür wurde in 38 Grundschulklassen der Unterricht in Deutsch, Kunst und Mathematik videogestützt beobachtet. Kooperationspartner sind die Universität Bamberg, die Universität Kassel und das DIPF.

70 Das Projekt wurde gemeinsam von der Humboldt-Universität Berlin und dem DIPF durchgeführt (Laufzeit: 2005–2007).

über das Fachportal Pädagogik für Re- und Sekundäranalysen zur Verfügung gestellt⁷¹. Aus Gründen des Persönlichkeitsschutzes betroffener Personen sind nicht alle Inhalte der Datenbank allgemein zugänglich. D.h. in diesem Fall, dass aufgrund der Rechtslage der Zugriff auf das Filmmaterial nur zu ausgewiesenen Forschungszwecken möglich ist. Die Legitimation ist mit geeigneten Unterlagen nachzuweisen. Der Zugang auf die Filme in den Formaten MPEG, AVI oder MWV ist zudem passwortgeschützt. Neben den zu Zeiten der DDR entstandenen Unterrichtsfilmen der Humboldt-Universität sind seit Oktober 2010 auch Filme der ehemaligen Akademie der Pädagogischen Wissenschaften der DDR in Berlin sowie der Universität Potsdam über das Fachportal Pädagogik verfügbar und es werden kontinuierlich neue Filme eingestellt. Derzeit liegt das Datenvolumen bei ca. 600 Gigabyte.

Außer auf die Videografien der Unterrichtsqualitätsforschung können Bildungsforscher auch auf qualitative Forschungsdaten für den Bereich der pädagogischen Praxis und der Rekonstruktion von Lehr-Lernprozessen zugreifen. Das Archiv für pädagogische Kasuistik (ApaeK)⁷² – angesiedelt am Institut für die Pädagogik der Sekundarstufe der Goethe-Universität Frankfurt am Main – verwaltet eine Online-Archivdatenbank mit derzeit 1414⁷³ teilweise zugangsbeschränkten Datensätzen.⁷⁴ Die pädagogische Kasuistik definiert sich über ihre Methodik, Dokumente der pädagogischen Praxis, wie z.B. Unterrichtsprotokolle (anonymisierte Transkripte) oder Unterrichtsplanungen als Einzelfall zu untersuchen und im Sinne verallgemeinerbarer Ableitungen zu analysieren.⁷⁵ Neben dem Fallmaterial werden auch die dazugehörigen Analysen online veröffentlicht, die Dokumente liegen zumeist im PDF-Format vor.

Die Bereitstellung dieser Art von Forschungsdaten konzentriert sich neben dem ApaeK auf zwei weitere Stellen in Deutschland. Dokumente der pädagogischen Kasuistik sind auch über das Fallarchiv Schulpädago-

71 Vgl. Fachportal Pädagogik (2011a).

72 Vgl. ApaeK 82011).

73 Stand: 14. April 2011.

74 Diese Dokumente sind nicht ausschließlich als Forschungsdaten zu interpretieren, sondern auch als Handreichungen für die pädagogische Praxis bzw. Unterrichtszwecke. Diese Daten sind wiederum sowohl qualitativ als auch quantitativ auswertbar.

75 ApaeK (2011).

gik⁷⁶ der Universität Kassel sowie die Arbeitsstelle Pädagogische Kasuistik⁷⁷ der TU Berlin verfügbar. Das Archiv der TU Berlin mit Protokollen teilnehmender Beobachtung, auditiven und audio-visuellen Dokumentationen und deren Transkription befindet sich derzeit noch im Aufbau. Diese drei Beispiele der Bereitstellung untermauern die Tendenz zur Dezentralität des Zugangs auf qualitative Forschungsdaten in der Bildungsforschung.

Zugang zu älteren Forschungsdaten und Retrodigitalisaten der Bildungsforschung

Insbesondere für die Historische Bildungsforschung ist der Zugriff auf ältere Forschungsdaten, wie z.B. Zeitungsartikel, Tagebücher, Klassenbücher oder andere Unterrichtsmaterialien von Bedeutung. Durch die Quellenorientierung des Forschungsbereichs ist der Zugang auf Retrodigitalisate unabdingbar; dieser erfolgt z.B. über die Bibliothek für Bildungsgeschichtliche Forschung (BBF) mit ihren Services *Scripta Paedagogica Online*⁷⁸, *Pictura Paedagogica Online*⁷⁹ und *Vox Paedagogica Online*⁸⁰. Inhaltliche Überschneidungen gibt es naturgemäß mit der Geschichtswissenschaft. Ein Beispiel für die Infrastrukturvernetzung von Historischer Bildungsforschung und Geschichtswissenschaft ist die Online-Publikationsplattform *perspectivia.net*⁸¹. Hier werden Retrodigitalisate von Publikationen der deutschen geisteswissenschaftlichen Auslandsinstitute und der Partnerorganisationen wie z.B. Zeitschriften, Rezensionen, Monografien und Tagungsdokumentationen frei zugänglich veröffentlicht.

Retrodigitalisate in der Bildungsforschung sind nicht zwangsläufig „ältere“ Daten. Auch im Rahmen von entwicklungsorientierten Forschungsansätzen, wie z.B. bei den Längsschnitt- und Panelstudien der Bildungsforschung, werden Retrodigitalisate als Kontextinformationen zeitnah dokumentiert. Beispiele hierfür sind die im Rahmen der Pythagoras-Studie digitalisierten Arbeitsblätter bzw. Unterrichtsmaterialien. Auch diese Art

76 Vgl. online-Fallarchiv (2011).

77 Vgl. TU Berlin (2011).

78 Das geschätzte Datenvolumen der Scripta beträgt 3,4 Terrabyte (vgl. BBF (2011a)).

79 Vgl. ppo (2011).

80 Vgl. BBF (2011b).

81 Vgl. perspectivia.net (2011).

retrodigitalisierter Daten soll im Zuge des DIPF-Projekts „Audiovisuelle Medien für die Unterrichtsforschung“ zur wissenschaftlichen Nachnutzung veröffentlicht und langzeitarchiviert werden.

Diese Überlegung macht auch deutlich, dass – insbesondere im Kontext qualitativer Ansätze – zwischen zwei Typen von Forschungsdaten in der Bildungsforschung (und ggf. auch in verwandten Forschungsbereichen) unterschieden werden muss: 1) Digitalisierte Daten als *Ergebnisse von Forschung* (z.B. Unterrichts-Videos, die auf der Basis eines konkreten Forschungsplans entstehen) und 2) digitalisierte Daten als *Gegenstände von Forschung* („Rohdaten“ wie z.B. Fotografien oder Zeichnungen, die als Abbild historischer Bildungsprozesse ohne Intention einer wissenschaftlichen Untersuchung entstanden sind).

Metadatenstandards für die Domäne der Bildungsforschung

Bisher hat sich für die Domäne der Bildungsforschung noch kein einheitlicher Metadatenstandard zur Dokumentation und zum Austausch von Forschungsdaten durchgesetzt, der die Bandbreite der heterogenen Datenbestände vollständig abbildet. Für einen Teilbereich der quantitativen Bildungsforschung – insbesondere im Bereich der großen Schulleistungs- und Lebenslaufstudien (PISA, NEPS) mit einer Vielzahl an statistischen Daten – kann jedoch aufgrund der inhaltlichen und methodischen Überschneidungen mit den Sozialwissenschaften die Metadatenspezifikation DDI⁸² als Ausgangspunkt für die Etablierung eines Metadatenstandards in der Domäne gelten.

Mit DDI lassen sich z.B. Namen und Werte von Variablen, Sampling-Methoden, die geografische Abdeckung und Zugangsbedingungen von Studien maschinenlesbar modellieren. Die aktuelle Version der Spezifikation, DDI 3.0, berücksichtigt zudem den gesamten Lebenszyklus von Daten innerhalb einer Studie, u.a. Informationen über die Finanzierung und die Archivierung der Daten sowie die Datenanalyse.⁸³ Zudem ermöglicht DDI das Mapping auf Dublin Core und MARC. DDI konzentriert sich bisher primär auf die

82 DDI (Data Documentation Initiative) ist eine XML-basierte Metadatenspezifikation, die die Datendokumentation und den Austausch von Forschungsdatensätzen im Bereich der Sozialwissenschaften ermöglicht (vgl. DDI (2011)).

83 Vgl. Vardigan; Heus; Thomas (2008), S. 110.

Beschreibung quantitativer Daten, jedoch gibt es im Rahmen der DDI-Initiative bereits erste Ansätze zur Integration qualitativer Befunde.⁸⁴

DDI wird bereits von einigen namhaften Datenhaltern, deren Scope Bildungsdaten umfasst, verwendet.⁸⁵ Hier bestehen jedoch – bedingt durch unterschiedliche Benennungen oder abweichende Belegungen von DDI-Feldern – noch Heterogenitäten in Bezug auf die Metadatenauszeichnung, die über Mapping-Routinen nur unzureichend aufgefangen werden können. Aus diesem Grund wird derzeit am IQB – zusammen mit mehreren Partnern⁸⁶ – an einer Weiterentwicklung des DDI-Standards für den Bereich der Bildungsforschung gearbeitet.

Als kompatible Untermenge zu DDI 3.0 soll *EduDDI* als Metadatenvokabular für die speziellen Bedürfnisse der Bildungsforschung entwickelt werden. Eine mögliche Erweiterung betrifft z.B. die Modellierung von Testskalen für Schulleistungsstudien. Ein wichtiger Aspekt, der von DDI ebenfalls nicht abgebildet wird, ist die Klassifizierung der Güte von Items⁸⁷ dieser Skalen zur Kompetenzmessung. Die Dokumentation der in *EduDDI* modellierten Studien soll möglichst zeitnah erfolgen und durch die einheitliche Struktur der Metadaten Vergleiche zu anderen Studien ermöglichen. Zudem soll durch *EduDDI* das Bewusstsein für die Notwendigkeit der Datenübergabe geschärft werden. Derzeit befindet sich *EduDDI* noch in der Entwicklungsphase.⁸⁸

Datenschutzaspekte

In der erziehungswissenschaftlichen Bildungsforschung werden häufig persönlichkeitsbezogene Daten erhoben, z.B. zu Lebensläufen, Bildungs-

84 Vgl. Corti; Gregory (2010).

85 Hier sind IQB, BIBB, GESIS, IAB, IZA oder ZPID (Leibniz-Zentrum für Psychologische Information und Dokumentation) zu nennen.

86 Weitere Projektpartner sind das DIPF, das TestDaF-Institut, das *IEA Data Processing and Research Center*, die Universität Bamberg (NEPS) sowie GESIS.

87 Die Items können je nach Schwierigkeit und Verlässlichkeit klassifiziert werden.

88 *EduDDI* ist an das Projekt NEPS angebunden, d.h. die LZA der dort gewonnenen Daten soll zeitgleich mit der Erhebung und Auswertung im *EduDDI*-Format erfolgen. Ein *Item- und Metadatenbuilder* soll zu einem generischen Instrument für die DDI-basierte und kollaborative Datendokumentation und Variablengenerierung weiterentwickelt werden. Der Schwerpunkt des Projekts liegt jedoch in der standardkonformen Konzeption des Tools und in der Entwicklung einer komfortablen und konsistenten Nutzerschnittstelle. Für weitere Informationen zu *EduDDI* vgl. Mechtel; Knigge (2009).

abschließen oder bildungsbezogenen Entscheidungsprozessen. Zudem werden Audio- und Videoaufnahmen als Erhebungsinstrumente qualitativer Forschung eingesetzt, um Kinder und Erwachsene in konkreten Verhaltenssituationen zu beobachten und Bildungsszenarien zu dokumentieren. In diesen Fällen sind z.B. die Rechte auf Wahrung des Persönlichkeitsschutzes (etwa Rechte am eigenen Wort und Bild) betroffen, „da die Daten nicht wie bei anderen Datenarten anonymisiert werden können“⁸⁹ bzw. dies nur mit hohem organisatorischem Aufwand machbar ist. Hierdurch treten insbesondere dann spezifische datenschutzrechtliche und forschungsethische Probleme auf,

„[...] wenn Kinder oder Personen erforscht werden, die nicht in der Lage sind, ihre Rechte selbst angemessen wahrzunehmen und zu vertreten. Datenschutzrechtlich relevante Fragen können auch dann auftreten, wenn es in historischer bzw. zeitgeschichtlicher Forschung um die konkrete Darstellung einzelner Institutionen oder Personen geht.“⁹⁰

Die Intention des Gesetzgebers in Bezug auf die Wahrung der Persönlichkeitsrechte wird durch die sogenannte *Zweckbindungsklausel* des Bundesdatenschutzgesetzes deutlich: Danach ist die Übermittlung personenbezogener Daten ohne Einwilligung des Betroffenen zulässig, wenn

„[...] es zur Durchführung wissenschaftlicher Forschung erforderlich ist, das wissenschaftliche Interesse an der Durchführung des Forschungsvorhabens das Interesse des Betroffenen an dem Ausschluss der Zweckänderung erheblich überwiegt und der Zweck der Forschung auf andere Weise nicht oder nur mit unverhältnismäßigem Aufwand erreicht werden kann.“⁹¹

Auch in den Datenschutzgesetzen der Bundesländer finden sich diese gesetzlichen Erlaubnistatbestände⁹². Dies bedeutet, dass

„[...] neue Objekte ohne Personendatenbezug (z.B. Statistiken) einer breiten Öffentlichkeit zugänglich gemacht werden können, während Materialien, die eine Identifikation der Betroffenen erlauben, Zugangsrestriktionen zu

89 Hugener; Pauli; Reusser (2006), S. 39.

90 DGfE (2006), S. 33.

91 §14 Abs. 2 Ziffer 9 BDSG.

92 Im Land Berlin ist z.B. die Übermittlung personenbezogener Daten ohne Einwilligung des Betroffenen zulässig, wenn dies nur für „bestimmte Forschungsarbeiten“ geschieht und „wenn das öffentliche Interesse an der Durchführung des Forschungsvorhabens die schutzwürdigen Belange des Betroffenen erheblich überwiegt und der Zweck der Forschung nicht auf andere Weise erreicht werden kann“ (§ 30 Abs. 1 Ziffer 2 BlnDSG).

unterworfen sind, die eine Verwendung lediglich für wissenschaftliche Forschungszwecke sicherstellen.“⁹³

Die Zweckbindungsklausel gilt explizit nicht für den Einsatz von audiovisuellen Rohdaten – beispielsweise in der Form von Unterrichtsvideos – zu Lehr- und Lernzwecken.

Neben den rechtlichen Vorgaben des Bundes und der Länder wurden für die Erziehungswissenschaft bereits Empfehlungen zur Anonymisierung von qualitativen Daten durch die Deutsche Gesellschaft für Erziehungswissenschaft (DGfE) vorgelegt.⁹⁴ Rechtlichen Problemen im Rahmen der *Biografieforschung*, der *Fotoanalyse*, der *Videografie*, der *Evaluationsforschung* und der *Fallgeschichte einzelner Institutionen* soll durch folgende Datenschutzmaßnahmen begegnet werden: 1) Absicherung aller am Forschungsprozess Beteiligten durch schriftliche Verträge mit Beschreibung von Zielen, Methoden, Verlauf und Publikationsabsicht sowie Zustimmung zum Gebrauch der Daten, 2) Sparsamkeitsgebot für Publikationen, d.h. es sollten möglichst wenige konkrete Fallbeispiele beschrieben werden, 3) abstrakte Formulierungssprache bei der Publikation von Daten und Ergebnissen und 4) Einschaltung der jeweils zuständigen Datenschutzbeauftragten.⁹⁵

Für die Veröffentlichung von personenbezogenen Forschungsprimärdaten – beispielsweise von audiovisuellen Rohdaten – ist es notwendig, ein hohes Maß an Rechtssicherheit zu gewährleisten. Entsprechend müssen technisch-organisatorische Zugriffskonzepte für das Rechtemanagement inklusive Zugangsbeschränkungen entwickelt werden. Im Rahmen des Projekts „Audiovisuelle Videodaten in der Unterrichtsforschung“ wurde dies auf der Basis einer in Auftrag gegebenen Rechtsexpertise⁹⁶ realisiert. In diesem Kontext wurden bspw. die im Projekt erteilten Einwilligungen von Lehrern, Schülern und Eltern geprüft und datenschutzrechtliche Implikationen zur Freiwilligkeit dieser Einwilligungen und zum Widerrufsrecht abgeleitet.

Auch für das Projekt „Datenbank Schulqualität“ – für das keine personenbezogenen Daten veröffentlicht werden – wurde ein Rechtsgutachten

93 Goebel; Scheller (2009).

94 Vgl. DGfE (2006), S. 33f.

95 Vgl. ebenda.

96 Vgl. ebenda.

in Auftrag gegeben. Relevante Themen sind dabei urheber- und markenrechtliche Aspekte sowie Datenschutzaspekte in Bezug auf die Analyse des Nutzerverhaltens in der Datenbank (Logfile-Analyse) und die persönliche Registrierung von Nutzern zur Zusammenstellung eigener Fragebogen. Insbesondere bei der Veröffentlichung von Skalen, die teilweise aus Publikationen übernommen oder durch Verlage lizenziert sind, muss die Einwilligung zur Veröffentlichung vorliegen bzw. müssen die entsprechenden Dokumente durch Nutzungsrestriktionen gekennzeichnet werden.⁹⁷

6.4 Interne Organisation

Bisher wird die LZA von Forschungsdaten in der Bildungsforschung noch nicht nach festen Regeln organisiert. Dennoch werden im Forschungsbereich bereits – exemplarisch für fest definierte Datenbestände – Standards und Prozesse für die LZA entwickelt wie z.B. im Projekt „Audiovisuelle Videodaten in der Unterrichtsforschung“ des DIPF (s. Kap. 6.3). Auch wenn diese Prozesse bisher noch nicht in der Form institutionalisiert sind, wie dies z.B. im Bereich des Open Access von Text-Publikationen in der Erziehungswissenschaft der Fall ist, so lassen sich hier erste Anforderungen für *Good Practices* des Forschungsdatenmanagements ableiten.

Bisher beschäftigen sich nur wenige Mitarbeiter der Informations- und Dokumentationsinfrastruktur in der Bildungsforschung mit der Verfügbarmachung und der LZA von Forschungsdaten. Eine Professionalisierung in Richtung eines elaborierten Objekt- und Metadatenmanagements inklusive spezifischem Know-how über Forschungsdaten in der Domäne steht noch aus. Ein erster Schritt ist die Teilnahme an projektunterstützten Weiterbildungsmaßnahmen zur LZA (z.B. die nestor *Spring School*) bzw. zum Metadatenmanagement (z.B. EduDDI-Workshops im Rahmen von NEPS).

Auch die Archivierungskapazitäten der involvierten Institutionen sind eine wichtige Voraussetzung für die Professionalisierung in der Domäne. Bisher werden für die LZA von Forschungsdaten ausschließlich die Services externer Dienstleister und Kompetenzzentren aus dem bibliothe-

97 Vgl. Goebel; Scheller (2011).

karischen Umfeld in Anspruch genommen (s. Projekt DP4lib Kap. 6.2). Die hohen Hardware- und Softwareanforderungen an den Betrieb eines Rechenzentrums für die LZA von Forschungsdaten sind von kleinen bis mittleren Einrichtungen meist nicht zu leisten.

6.5 Perspektiven und Visionen

Nicht nur im Bereich der Bildungsforschung herrschen noch große Unsicherheiten in Bezug auf Rechtssicherheit und „ethische Sicherheit“ beim Umgang mit Forschungsdaten. Analog zu den DFG-Empfehlungen zu Open Access und Urheberrecht⁹⁸ sind daher übergreifende Ethikregeln für die Benutzung und Verarbeitung von Forschungsdaten notwendig. Hier bietet sich die Einbeziehung der Fachgesellschaften zur Erarbeitung spezifischer Datenschutzempfehlungen an. Ein mögliches Beispiel ist die Erweiterung der DGfE-Empfehlung für den Umgang mit qualitativen Daten um Aspekte des Urheberrechts und der Persönlichkeitsrechte sowie des Datenmanagements zur Vereinfachung von Sekundäranalysen (s. Datenschutzaspekte im Kap. 6.3) – auch um den nachträglichen Zugriff auf die Forschungsdaten durch die LZA zu ermöglichen. Neben institutsübergreifenden Regeln ist die Erarbeitung von jeweils institutsinternen Regeln – in Abstimmung mit den Datenproduzenten – eine maßgebliche Anforderung, um der Vielfalt der Datenarten in der Bildungsforschung gerecht werden zu können.

Die Wissenschaft respektive die Erziehungswissenschaft sollte stärker in die strategischen informationsinfrastrukturellen Anforderungen einbezogen werden. Eine empirische Untersuchung zu den konkreten Bedürfnissen und Problembereichen des Zugangs zu Forschungsdaten, wie sie bereits für qualitative Daten in den Sozialwissenschaften erarbeitet wurde⁹⁹, steht für die Bildungsforschung noch aus. Kollaborative Projekte, bei der Bildungsforscher und Informationsspezialisten bei der Datendokumentation und Metadatenentwicklung zusammenarbeiten (NEPS), dienen zudem als *use case* mit Ausstrahlungseffekt auf den ganzen For-

98 Hier sind zwei Positionspapiere der DFG relevant: Open Access und Urheberrecht (vgl. DFG (2006); DFG (2009d)).

99 Vgl. Medjedović; Witzel (2010).

schungsbereich. Insbesondere für den Bereich der Metadatenstandards gibt es eine kooperative Herausforderung, den vorhandenen Standard DDI in Hinblick auf die Beschreibung qualitativer Datenarten weiterzuentwickeln.

Hinsichtlich der Verbindung von Datenschutz und Datenaufbereitung besteht auf Seiten der Bildungsforscher hoher Informationsbedarf. Hier sind zum einen Transferleistungen der Informations- und Dokumentationsdienstleister gefragt, wie z.B. die Aufbereitung von Daten, um sie für Sekundäranalysen oder als konkrete Planungshilfe im Forschungsprozess nutzbar zu machen. Zum anderen muss die Gestaltung der Prozesse unter Einbindung der Wissenschaftler als Datenproduzenten bzw. Nutzer der Datenrepositorien erfolgen, um fachspezifische Anforderungen zu berücksichtigen. Besonders relevant ist in diesem Zusammenhang auch die Implementierung von Strategien der *Data Curation*, die eine prozessorientierte Sicht des Forschungsdatenmanagements beinhalten und bereits vor der Erhebung der Daten deren Pflege und organisierten Umgang mit berücksichtigt.¹⁰⁰ Dies könnte den Weg ebnen für einen bedarfsgerechten und serviceorientierten Zugang auf Forschungsdaten in der Domäne.

Vertrauenswürdigkeit von Datenarchiven ist eine essentielle Forderung der Wissenschaft. Vor diesem Hintergrund sollte ein Zertifizierungsverfahren für „vertrauenswürdige Forschungsdatenzentren“ angestrebt werden. Den Maßstab in Bezug auf die Umsetzung einer solchen qualitätssichernden Strategie der Forschungsdatenspeicherung stellen die DINI-zertifizierten Repositorien dar. Ein Bestandteil des Qualitätsmanagements ist etwa die Vergabe von *persistent identifiers*, um die Identifikation und Zitation von Forschungsdaten sicherzustellen. Die Zertifizierung von Forschungsdatenzentren für die Bildungsforschung verlangt dabei spezifische Anforderungen, z.B. an den Umgang mit großen Mengen an qualitativen Daten oder die Wahrung des Persönlichkeitsschutzes.

Gemäß dem interdisziplinären Charakter der Bildungsforschung besteht eine zentrale Vision darin, eine adäquate interdisziplinäre Infrastruktur für Forschungsdaten zu entwickeln, die über die einzelnen

100 Vgl. Pampel; Bertelmann; Hobohm (2011).

Forschungsdatenzentren hinweg ein hohes und gesichertes Niveau an Vertrauenswürdigkeit, Interoperabilität und Services für den gesamten Datenraum der Fächergruppe gewährleistet.

Vor diesem Hintergrund ist im Bereich der Forschungsdatenzentren ein noch stärker koordiniertes Vorgehen wünschenswert, welches unter dem Vorzeichen der Interdisziplinarität tragfähige LZA-Strategien abstimmt und weiterentwickelt. Gerade die projektbasierte Bildungsforschung in den kleinen und mittleren Einrichtungen benötigt eine übergreifende „Anlaufstelle“, die relevante Bestände der Domäne im Sinne der LZA bündelt, harmonisiert und sichert. Neben der Langzeitsicherung sind die Langzeitverfügbarkeit und entsprechende Zugriffsservices wesentliche Aspekte einer wissenschaftsorientierten Forschungsdatenstrategie. Ein zentrales Datenservicezentrum für Forschungsdaten der Erziehungswissenschaft – auch in Hinblick auf die internationale Anschlussfähigkeit der Disziplin – ist daher erstrebenswert.